AMERICAN SOCIETY FOR MICROBIOLOGY | Antimicrobial Agents and Chemotherapy®

EPIDEMIOLOGY AND SURVEILLANCE

Genomic Epidemiology of Complex, Multispecies, Plasmid-Borne *bla*$_{KPC}$ Carbapenemase in *Enterobacterales* in the United Kingdom from 2009 to 2014

Nicole Stoesser,[a,b] Hang T. T. Phan,[b,c] Anna C. Seale,[a,d] Zoie Aiken,[e] Stephanie Thomas,[e] Matthew Smith,[e] David Wyllie,[a] Ryan George,[e] Robert Sebra,[f] Amy J. Mathers,[g,h] Alison Vaughan,[a] Timothy E. A. Peto,[a,b,i] Matthew J. Ellington,[j] Katie L. Hopkins,[j] Derrick W. Crook,[a,b,i] Alex Orlek,[a,b] William Welfare,[k] Julie Cawthorne,[e] Cheryl Lenney,[e] Andrew Dodgson,[e,l] Neil Woodford,[j] A. Sarah Walker,[a,b,i] the TRACE Investigators' Group

[a]Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom
[b]NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, University of Oxford, Oxford, United Kingdom
[c]Faculty of Medicine, University of Southampton, Southampton, United Kingdom
[d]Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom
[e]Manchester University NHS Foundation Trust, Manchester, United Kingdom
[f]Icahn Institute and Department of Genetics and Genomic Sciences, Icahn School of Medicine, Mount Sinai, New York, USA
[g]Division of Infectious Diseases and International Health, Department of Medicine, University of Virginia Health System, Charlottesville, Virginia, USA
[h]Clinical Microbiology Laboratory, Department of Pathology, University of Virginia Health System, Charlottesville, Virginia, USA
[i]NIHR Biomedical Research Centre, University of Oxford, Oxford, United Kingdom
[j]Antimicrobial Resistance and Healthcare Associated Infections (AMRHAI) Reference Unit, National Infection Service, Public Health England, London, United Kingdom
[k]Public Health England North West, Manchester, United Kingdom
[l]Public Health Laboratory, Public Heath England, Manchester, United Kingdom

Nicole Stoesser and Hang T. T. Phan contributed equally to this work. Nicole Stoesser coordinated the study and, therefore, is listed first.

ABSTRACT Carbapenem resistance in *Enterobacterales* is a public health threat. *Klebsiella pneumoniae* carbapenemase (encoded by alleles of the *bla*$_{KPC}$ family) is one of the most common transmissible carbapenem resistance mechanisms worldwide. The dissemination of *bla*$_{KPC}$ historically has been associated with distinct *K. pneumoniae* lineages (clonal group 258 [CG258]), a particular plasmid family (pKpQIL), and a composite transposon (Tn*4401*). In the United Kingdom, *bla*$_{KPC}$ has represented a large-scale, persistent management challenge for some hospitals, particularly in North West England. The dissemination of *bla*$_{KPC}$ has evolved to be polyclonal and polyspecies, but the genetic mechanisms underpinning this evolution have not been elucidated in detail; this study used short-read whole-genome sequencing of 604 *bla*$_{KPC}$-positive isolates (Illumina) and long-read assembly (PacBio)/polishing (Illumina) of 21 isolates for characterization. We observed the dissemination of *bla*$_{KPC}$ (predominantly *bla*$_{KPC-2}$; 573/604 [95%] isolates) across eight species and more than 100 known sequence types. Although there was some variation at the transposon level (mostly Tn*4401a*, 584/604 [97%] isolates; predominantly with ATTGA-ATTGA target site duplications, 465/604 [77%] isolates), *bla*$_{KPC}$ spread appears to have been supported by highly fluid, modular exchange of larger genetic segments among plasmid populations dominated by IncFIB (580/604 isolates), IncFII (545/604 isolates), and IncR (252/604 isolates) replicons. The subset of reconstructed plasmid sequences (21 isolates, 77 plasmids) also highlighted modular exchange among non-*bla*$_{KPC}$ and *bla*$_{KPC}$ plasmids and the common presence of multiple replicons within *bla*$_{KPC}$ plasmid structures (>60%). The substantial genomic plasticity observed has important implications for our understanding of the epidemiology of transmissible carbapenem resistance in *Enterobacterales* for the implementation of adequate surveillance approaches and for control.

Citation Stoesser N, Phan HTT, Seale AC, Aiken Z, Thomas S, Smith M, Wyllie D, George R, Sebra R, Mathers AJ, Vaughan A, Peto TEA, Ellington MJ, Hopkins KL, Crook DW, Orlek A, Welfare W, Cawthorne J, Lenney C, Dodgson A, Woodford N, Walker AS, the TRACE Investigators' Group. 2020. Genomic epidemiology of complex, multispecies, plasmid-borne *bla*$_{KPC}$ carbapenemase in *Enterobacterales* in the United Kingdom from 2009 to 2014. Antimicrob Agents Chemother 64:e02244-19. https://doi.org/10.1128/AAC.02244-19.

Address correspondence to Nicole Stoesser, nicole.stoesser@ndm.ox.ac.uk.

Received 8 November 2019
Returned for modification 6 January 2020
Accepted 21 February 2020
Accepted manuscript posted online 24 February 2020
Published 21 April 2020

KEYWORDS *Enterobacterales*, KPC, whole-genome sequencing, short-read sequencing, long-read sequencing, carbapenemase, genomic epidemiology, outbreak analysis

Antimicrobial resistance (AMR) in *Enterobacterales* is a critical public health threat. Carbapenem resistance is of particular concern, and its evolution and spread in multiple species of *Enterobacterales* (i.e., carbapenemase-producing *Enterobacterales* [CPE]) is increasingly reported (1–4). The exchange of AMR genes, including carbapenem resistance genes, happens at multiple genetic levels (5) and is often facilitated by their presence on plasmids (circular DNA structures of variable size [2 kb to ~>1 Mb]) and/or other smaller mobile genetic elements (MGEs), such as transposons and insertion sequences (IS), that form part of the accessory genome.

Whole-genome sequencing (WGS) has significantly improved our understanding of infectious disease epidemiology and is used in both community-associated and nosocomial transmission analyses (6, 7). Although useful for delineating transmission routes in clonal, strain-based outbreaks, standard phylogenetic approaches and comparative analyses have been more difficult where multiple bacterial strains/species and transmissible resistance genes are involved (5). The reconstruction of the genetic structures of plasmids carrying relevant antimicrobial resistance genes using long-read sequencing has improved our understanding of the genetic complexity of the spread of important resistance genes but has been difficult to undertake on a large scale.

Although approximately 40 *Klebsiella pneumoniae* carbapenemase (KPC; encoded by bla_{KPC}) variants have been described (per NCBI's AMR reference gene catalogue, available at https://www.ncbi.nlm.nih.gov/pathogens/isolates#/refgene/), two have been most widely reported globally, namely, KPC-2 and KPC-3 (H272Y with respect to KPC-2; single-nucleotide difference in bla_{KPC} [C814T]) (8, 9). In the United Kingdom, the first KPC isolate identified was a KPC-4-containing *Enterobacter* sp. isolated in Scotland in 2003 (10), with subsequent identification of KPC-3 in isolates in the United Kingdom in 2007. From 2007, increasing numbers of suspected KPC isolates were referred to the Public Health England (PHE) Antimicrobial Resistance and Healthcare-Associated Infections (AMRHAI) Reference Unit, with the majority of confirmed KPC producers (>95%) being KPC-2 and from hospitals in North West England, first recognized in 2008 to 2009 (11). These isolates were predominantly bla_{KPC}-positive *Enterobacterales* cultured from patients in the Central Manchester University Hospitals NHS Foundation Trust (CMFT; now part of the Manchester University NHS Foundation Trust) (12). bla_{KPC} is thought to have been introduced into the region via a pKpQIL-like plasmid (13, 14), a plasmid backbone previously associated with the global dissemination of bla_{KPC} in *K. pneumoniae* clonal group 258 and already observed in other *K. pneumoniae* sequence types (STs) and species in an analysis of 44 UK KPC-*Enterobacterales* from 2008 to 2010 (14).

We used WGS to undertake a large-scale retrospective study of this multispecies, polyclonal dissemination of bla_{KPC} in major Manchester hospitals in northwest England from 2009, generating complete genome structures, including bla_{KPC} plasmids, for a subset of isolates. We contextualized our analysis of strains in Manchester by sequencing a subset of isolates from the local region (North West England) and other hospitals in the United Kingdom collected through a national bla_{KPC} surveillance program, with the goal of understanding the genetic structures associated with the regional emergence of bla_{KPC} in this setting.

RESULTS

Of 742 isolates identified for sequencing, 60 (8%) were not retrievable or cultivable from the laboratory archives. After deduplicating by taking the first bla_{KPC}-positive *Enterobacterales* (KPC-E) isolate per patient and excluding sequencing failures, any sequences without bla_{KPC} (assumed lost in culture), and mixtures (identified from

genomic data analysis; see Materials and Methods), 604 evaluable isolate sequences were included. These represented 327 archived isolates (54%) from inpatients in the early stages of the observed introduction of bla_{KPC} in the two Manchester hospitals studied (2009 to 2011), of which 309 and 18 isolates were from CMFT and the University Hospital of South Manchester NHS Foundation Trust (UHSM; now part of Manchester University NHS Foundation Trust), respectively; 78 (13%) later isolates from CMFT/UHSM (2012 to 2014); 119 (20%) isolates from other hospitals ($n = 15$ hospitals) in North West England (2009 to 2014, excluding CMFT and UHSM, up to the first 25 consecutive KPC-E isolates per hospital); 72 (12%) isolates from UK and Irish hospitals ($n = 72$ locations [$n = 4$ from Ireland]) outside the North West (2009 to 2014) (first KPC-E isolate per hospital); and 8 (1%) isolates from English outpatient/primary care settings (7 from the North West region, 1 from a southern UK location). The geographic and numerical distribution of isolates is depicted in Data S1 in the supplemental material.

Consistent with increasing numbers of bla_{KPC}-positive *Enterobacterales* reported nationally to the reference laboratory, cases in CMFT/UHSM also began to rise from 2009. Anecdotally the first cases were reported in 2009, with 63 carbapenem-resistant *Enterobacterales* cultured from 18,630 microbiological specimens processed (0.3%), with a 10-fold increase by 2014 (988/29593 [3%]) (Data S2).

bla_{KPC-2} dominates in the United Kingdom, but highly variable bla_{KPC} copy numbers and associated resistance marker profiles suggest a diverse, flexible accessory genome underpinning its spread. Although three bla_{KPC} variants were observed in the 604 included isolates, bla_{KPC-2} dominated ($n = 573$, 95%; bla_{KPC-3}, $n = 27$, 4%; bla_{KPC-4}, $n = 4$, 1%). Two isolates (0.3%; trace524 and trace534) showed evidence of mixed populations of bla_{KPC-2} and bla_{KPC-3}.

The median bla_{KPC} copy number estimate was 1.8 (interquartile range [IQR], 1.6 to 2.1), with a maximum of 8.2. bla_{KPC} copy number estimates strongly associated with meropenem MICs for the 588 isolates for which both were available (estimated +0.56 higher doubling dilution [95% confidence interval, +0.40, +0.72] per copy number, $P <$ 0.0001; Data S3) and also between approaches deriving estimates from short-read assemblies versus those from mapping to reconstructed genomes (Pearson's correlation coefficient, 0.97 [$P = 0.0001$]; Data S4) (see Materials and Methods). Across the three main species, bla_{KPC} copy numbers were higher in *K. pneumoniae* ($n = 525$ [87%], median of 1.8 [IQR, 1.6 to 2.1]) than in *Escherichia coli* (40 [7%], 1.7 [1.5 to 1.9]) or *Enterobacter cloacae* (26 [4%], 1.6 [1.4 to 2.0]) ($P = 0.0003$ by Kruskal-Wallis test) (Fig. 1A). Among common STs, copy numbers were highest in *K. pneumoniae* ST258 ($n = 65$ [11%], median of 2.4 [IQR, 1.8 to 2.9]) versus other species/STs ($n = 531$ [89%], median of 1.8 [1.6, 2.0]) ($P = 0.0001$ by Kruskal-Wallis test) (Fig. 1B and Data S5). Of note, bla_{KPC} copy number estimates represent an average across all individual cells sequenced; bla_{KPC} copy number estimates of <1 in a small number of isolates (20/604 [3%]) (median of 0.82 [IQR, 0.40 to 0.96]) suggest that a proportion of cells in the populations sequenced have lost their bla_{KPC}-harboring plasmid.

There were 364 distinct resistance marker profiles in isolates, with only 12% (74/604) of isolates sharing exactly the same profile as >10 other isolates (Data Set SD1). Other broad- or extended-spectrum beta-lactamase genes were also commonly present across isolates, including bla_{TEM} ($n = 452$, all bla_{TEM-1}), bla_{OXA} ($n = 492$; bla_{OXA-9} [$n = 425$] and bla_{OXA-1} [$n = 138$]), bla_{SHV} ($n = 497$), and bla_{CTX-M} ($n = 89$; $bla_{CTX-M-15}$ [$n = 57$] and $bla_{CTX-M-9}$ [$n = 28$]). Aminoglycoside resistance genes were also widely prevalent: *aac* ($n = 243$), *aph* ($n = 196$), and *ant* ($n = 373$). In terms of acquired quinolone resistance, 160 isolates contained *qnr* variants, and 137 isolates contained *aac(6')-Ib-cr*; no *qep* variants were seen.

bla_{KPC} in the United Kingdom is a multispecies, largely polyclonal phenomenon. In contrast to the almost uniform presence of bla_{KPC-2} in isolates, species and lineage diversity among our entire isolate collection was substantial, with eight different species among sequenced isolates. For species with developed multilocus sequence type (MLST) schemes, this represented a total of 102 different known species-ST combinations and 26 additional unknown species-ST combinations, namely, *K. pneu-*

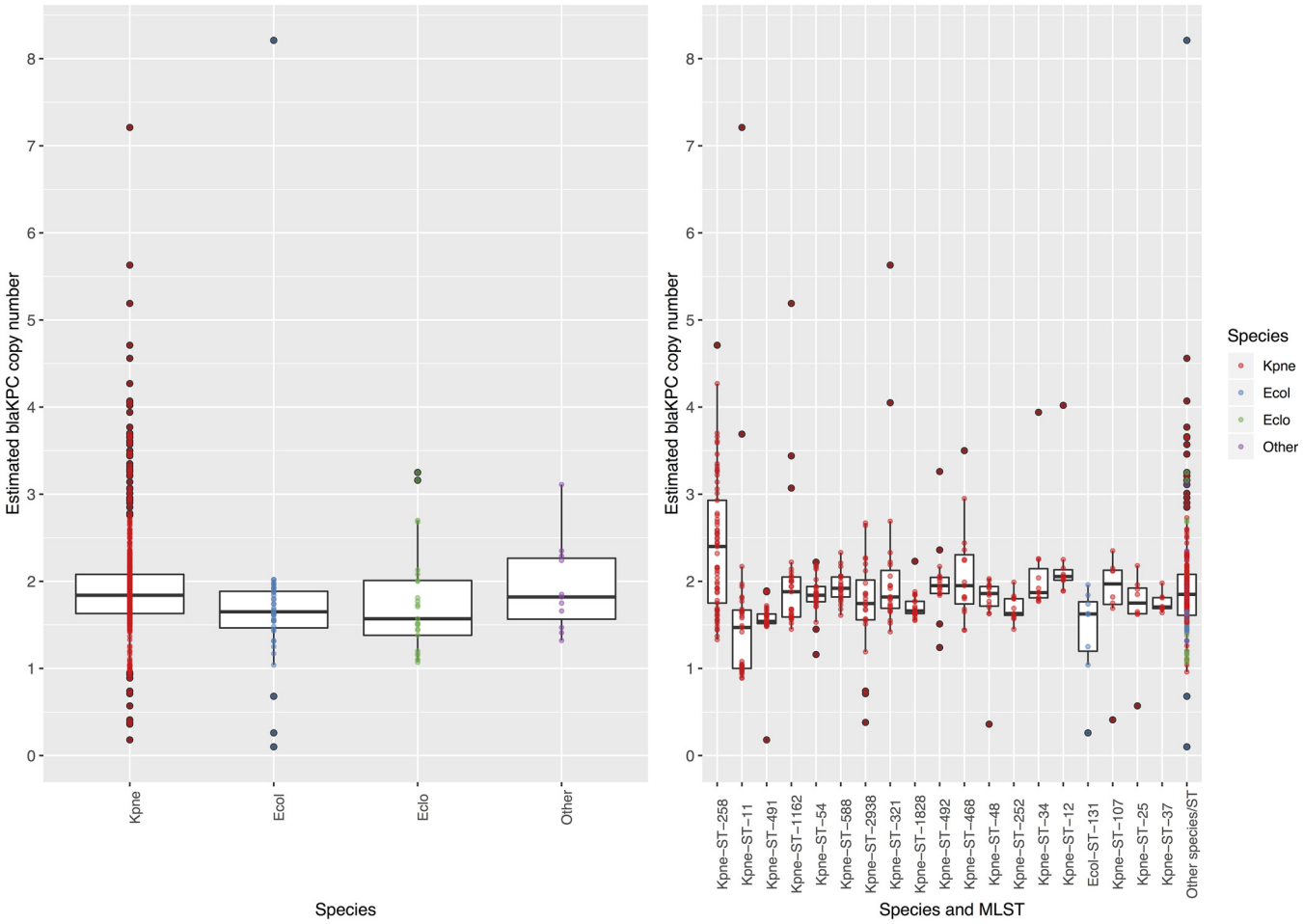

FIG 1 Estimated bla_{KPC} copy number distributions (derived from Illumina assemblies) within major species (left) and the top nineteen most common species/ST combinations (right) observed in the study. Other species were designated "other" (left) and other ST/species combinations as "other species/ST" (right). Dots represent estimated copy numbers for single isolates; boxplots represent median estimated bla_{KPC} copy numbers ± 1.58 × IQR/sqrt(*n*). Boxplots are ordered by most common species and species/ST categories, left to right, except for the other and other species/ST, assigned to the right of the plots. Kpne, *Klebsiella pneumoniae*; Ecol, *Escherichia coli*; Eclo, *Enterobacter cloacae*.

moniae ($n = 525$ isolates, 70 known STs, 20 unknown STs), *E. coli* ($n = 40$, 20 known STs, 1 unknown ST), *Enterobacter cloacae* ($n = 26$, 9 known STs, 2 unknown STs), *Klebsiella oxytoca* ($n = 6$, 3 known STs, 3 unknown STs), *Raoultella ornithinolytica* ($n = 4$), *Enterobacter aerogenes* ($n = 2$), *Serratia marcescens* ($n = 1$), and *Kluyvera ascorbata* ($n = 1$). The most common STs all were for *K. pneumoniae*, including ST258 ($n = 66$), ST11 ($n = 35$), ST491 ($n = 31$), ST1162 ($n = 29$), and ST54 ($n = 27$) (Fig. 2). Among these five common STs, the distribution of pairwise single-nucleotide variant (SNV) distances between isolates suggested that some isolates can be considered highly genetically related at the strain level; however, overall, when considering SNV distances alongside specific accessory genome differences, isolates from these STs were not clonal (Data S6 to S10). Notably, 42% (252/604) of isolates were found in STs represented by ≤10 isolates (Fig. 2).

Although some of the earliest sequenced isolates in the collection were KPC-*K. pneumoniae* ST258 and ST11 (both in 2009; two major KPC strains from CG258 circulating globally and in China at the time [8, 15]), and although KPC-producing *K. pneumoniae* ST258 appears to have been one of the earliest strains observed in the two Manchester hospitals (CMFT and UHSM), multiple diverse STs and species clearly were subsequently rapidly recruited specifically in these two hospital settings in 2010 and 2011, with bla_{KPC} emerging in at least 30 new species-ST groups per year (76% and 59%,

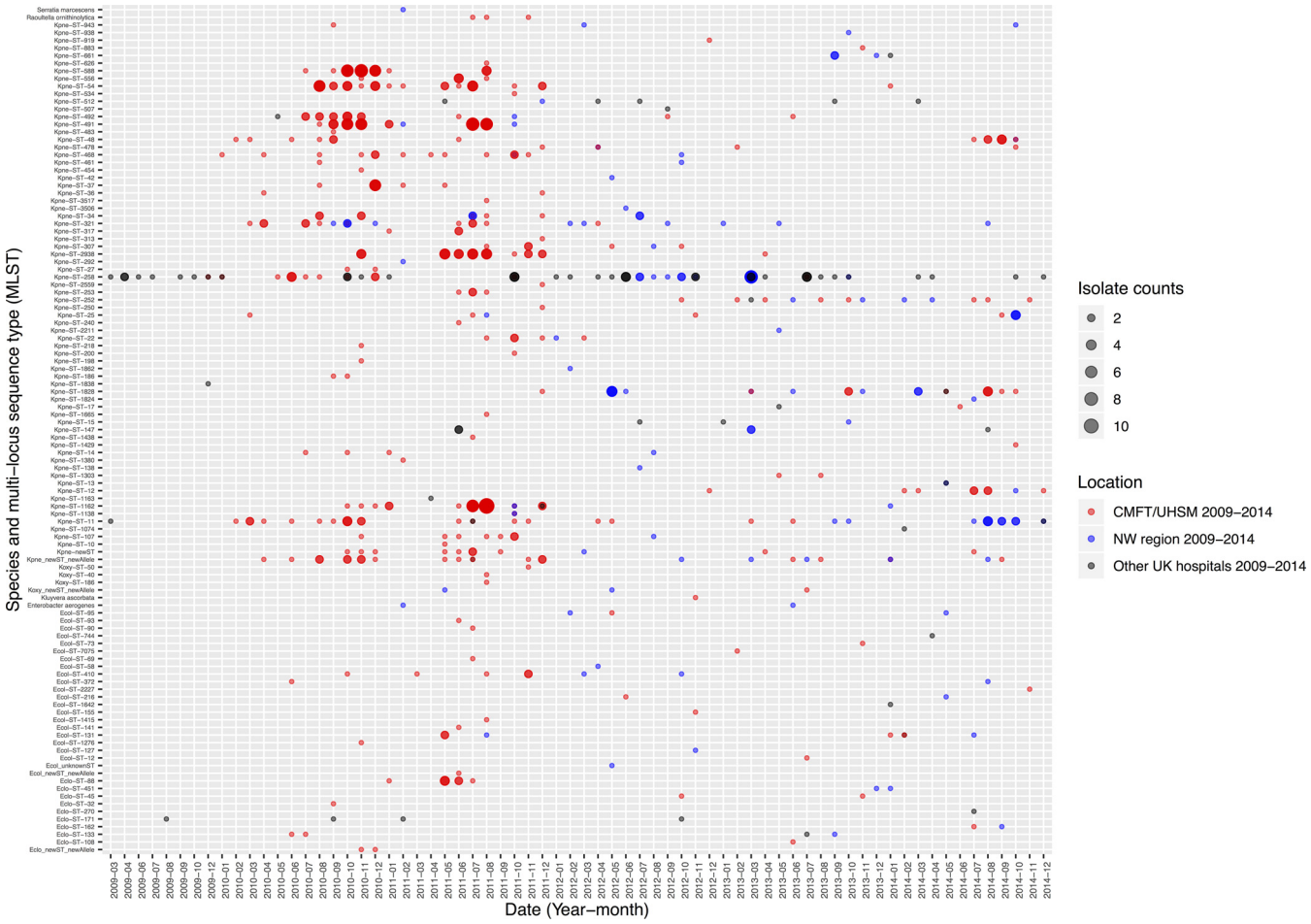

FIG 2 Incidence plot of bla_{KPC} species-ST by year-month and geography. Dots are colored by the location of isolate collection, as defined in Materials and Methods, and scaled by the number of isolates; the eight outpatient isolates were not plotted.

respectively, of first-per-patient carbapenem-resistant *Enterobacterales* culture positives sequenced) (Fig. 3).

Most bla_{KPC} members in the United Kingdom are supported by a conserved Tn*4401a* unit with uniform target site sequences, suggesting that direct Tn*4401* transposition is not the main mode of bla_{KPC} transmission in this context. In the absence of evidence of significant clonal spread by high-risk bacterial lineages, we explored the diversity among mobile genetic features. Tn*4401* is an ~10-kb transposon that has been the major transposable context for bla_{KPC} to date, and it is flanked by 5-bp signatures of transposition (target site duplications [TSD] or target site sequences [TSS]) (16, 17), with no known target site specificity (16). A predominant Tn*4401* isoform was associated with both bla_{KPC-2} and bla_{KPC-3} in this study, namely, Tn*4401a* (16), which occurred in 584/604 (97%) isolates (Fig. 4). Other known variants included Tn*4401b* ($n = 7$) and Tn*4401d* ($n = 3$). Only 20/584 (3%) isolates demonstrated evidence of SNV-level variation in Tn*4401a* (homozygous calls at 6 positions; heterozygous calls [i.e., mixed populations] at 3 positions).

bla_{KPC-2}-Tn*4401a* ($n = 539$ isolates) was predominantly flanked by the 5-bp TSD ATTGA, with 465/604 (77%; 465/539 [86%] of this subtype) isolates with this Tn*4401*/TSD combination throughout the study period (Fig. 4). In 74 other bla_{KPC-2}-Tn*4401a* isolates, Tn*4401a* was flanked by other TSS combinations, consistent with additional transposition events. Thirty-two of these were TSDs (16 AATAT-AATAT, 16 AGTTG-AGTTG), which have been described as being more consistent with interplasmid transposition of Tn*4401* (18), and 35 were nonduplicate TSS combinations (ATTGA with

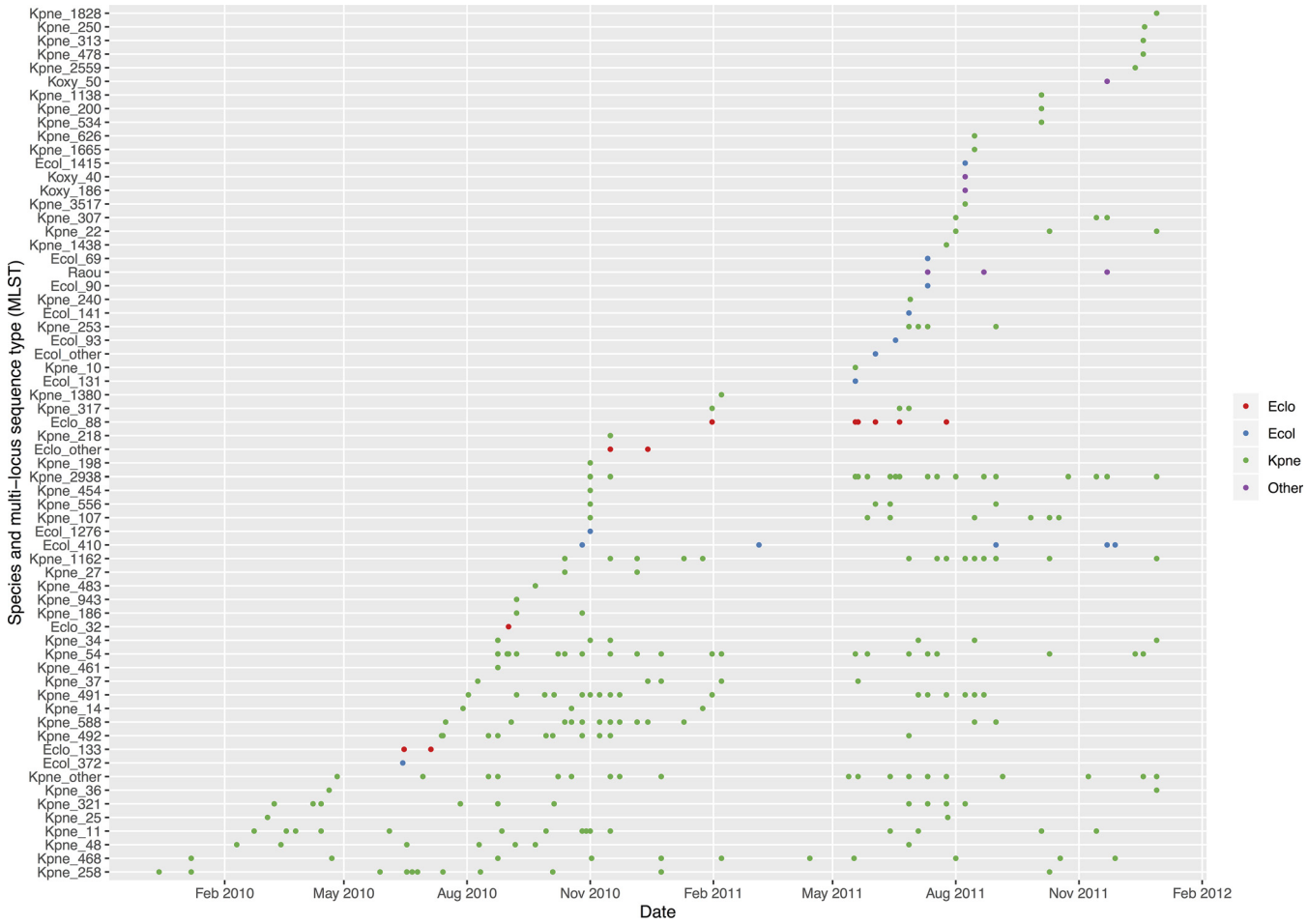

FIG 3 Incidence curve of species-ST in Manchester hospitals CMFT/UHSM, 2010 to 2012. Sequencing ascertainment of first-per-patient carbapenem-resistant *Enterobacterales* was 76% in 2010 and 59% in 2011.

either ATATA, TGGTA, CTGCC, AATAA, or AGGAT), described as more consistent with intraplasmid transposition. Evidence of multiple TSSs around bla_{KPC-2}-Tn*4401a* within single isolates was seen in 6 cases (i.e., multiple right and/or left Tn*4401* TSSs); 1 case had a right TSS present but no left TSS identified.

Plasmid replicon typing demonstrates diverse plasmid populations present in bla_{KPC}-positive isolates in the United Kingdom but with combinations of IncF, IncR, ColRNAI, and IncX3 replicons predominating. The 604 isolates contained 91 unique combinations of plasmid replicon family types, a crude proxy of plasmid populations present. However, it was not possible to determine the colocalization of specific replicon types on plasmid structures or direct associations with bla_{KPC} using this approach and short-read sequencing data. No isolate was replicon negative. However, there were seven predominant replicon family combinations (Fig. 5) represented in 443/604 (73%) isolates, and these included six major replicon family types, namely, IncF (FIB [found in $n = 580$ isolates] and FII [$n = 545$]), FIA ($n = 103$), IncR ($n = 252$), ColRNAI ($n = 86$), and IncX3 ($n = 60$). IncFIB/IncFII and IncFIB/IncFII/IncR populations were most widely distributed across species-STs (Data S11) and geographical regions as well as over time (Fig. 5). The diversity of plasmid backgrounds present in these isolates may facilitate opportunities for bla_{KPC} exchange among different plasmid families.

For many of the plasmid families, several different reference replicon sequences exist in the PlasmidFinder database, with a degree of homology among sequences in the same family, making it difficult to establish robustly which subtype of

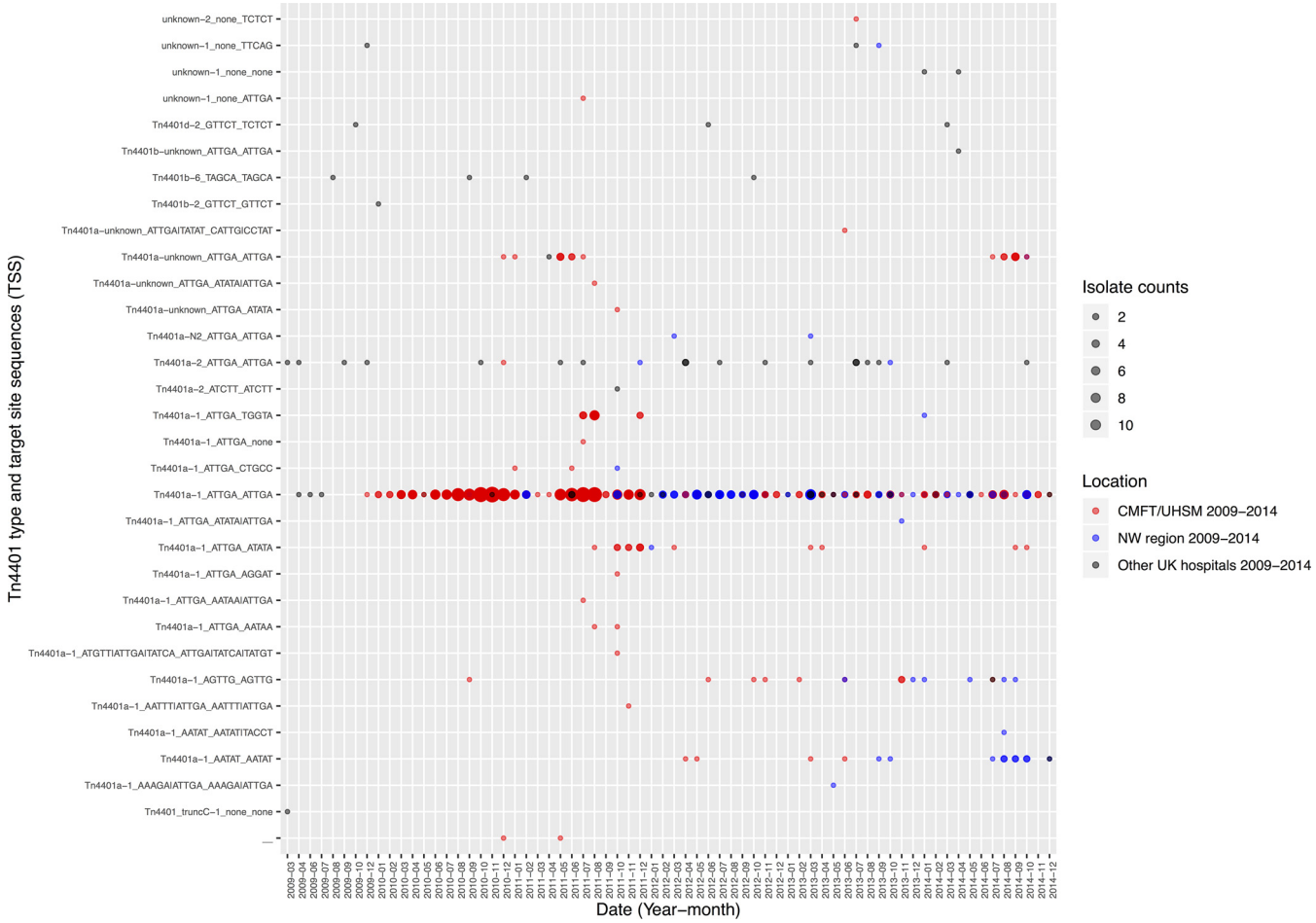

FIG 4 Incidence plot of Tn*4401* type and target site sequences by year-month and geography. Dots are colored by location of isolate collection, as defined in Materials and Methods, and scaled by the number of isolates; the eight outpatient isolates were not plotted.

replicon is present. However, restricting to 90% matches to reference replicon types (with matches being a composite of percent sequence identity times percent reference sequence coverage) for these common families, top matches found in more than 10% of isolates included IncFIB(K)_1_Kpn3_JN233704 (n = 490 isolates), IncFIB(pQil)_JN233705 (n = 299), IncR_1_DQ449578 (n = 253), IncFII(K)_1_CP000648 (n = 89; plasmid MLST $IncFII_{K1}$), ColRNAI_1_DQ298019 (n = 86), IncFIA(HI1)_1_HI1_AF250878 (n = 80), IncFII_1_pKP91_CP000966 (n = 90; plasmid MLST $IncFII_{K4}$), and IncX3_1_JN247852 (n = 61). At this more detailed typing resolution, we found 183 plasmid replicon subtype profiles among our isolates, only ten of which were shared across ≥10 isolates and two across more than 10% of isolates. The two most widespread profiles were IncFIB(pQil)_JN233705 + IncFIB(K)_1_Kpn3_JN233704 and IncFIB(K)_1_Kpn3_JN233704 + IncR_1_DQ449578 (Data S12).

Reference, short-read-based bla_{KPC} plasmid typing demonstrates that pKpQIL-like plasmids have been key in disseminating bla_{KPC} in the United Kingdom but that no known bla_{KPC} plasmid vector was responsible in almost a third of cases. We then tried to identify the most likely bla_{KPC}-associated plasmid candidate, as opposed to broadly profiling plasmid populations within isolates by plasmid replicon typing. Attempts to identify complete plasmids from short-read data by comparison to a reference plasmid database have been estimated as being correct in only ~45% to 85% of cases in previous studies (5, 19). However, 13/14 (93%) of isolates for which we had hybrid assemblies (see Materials and Methods) with only one completely reconstructed bla_{KPC} plasmid had the correct top match using this bla_{KPC} plasmid typing method (Data Set SD2). We therefore compared all short-read sequences with our

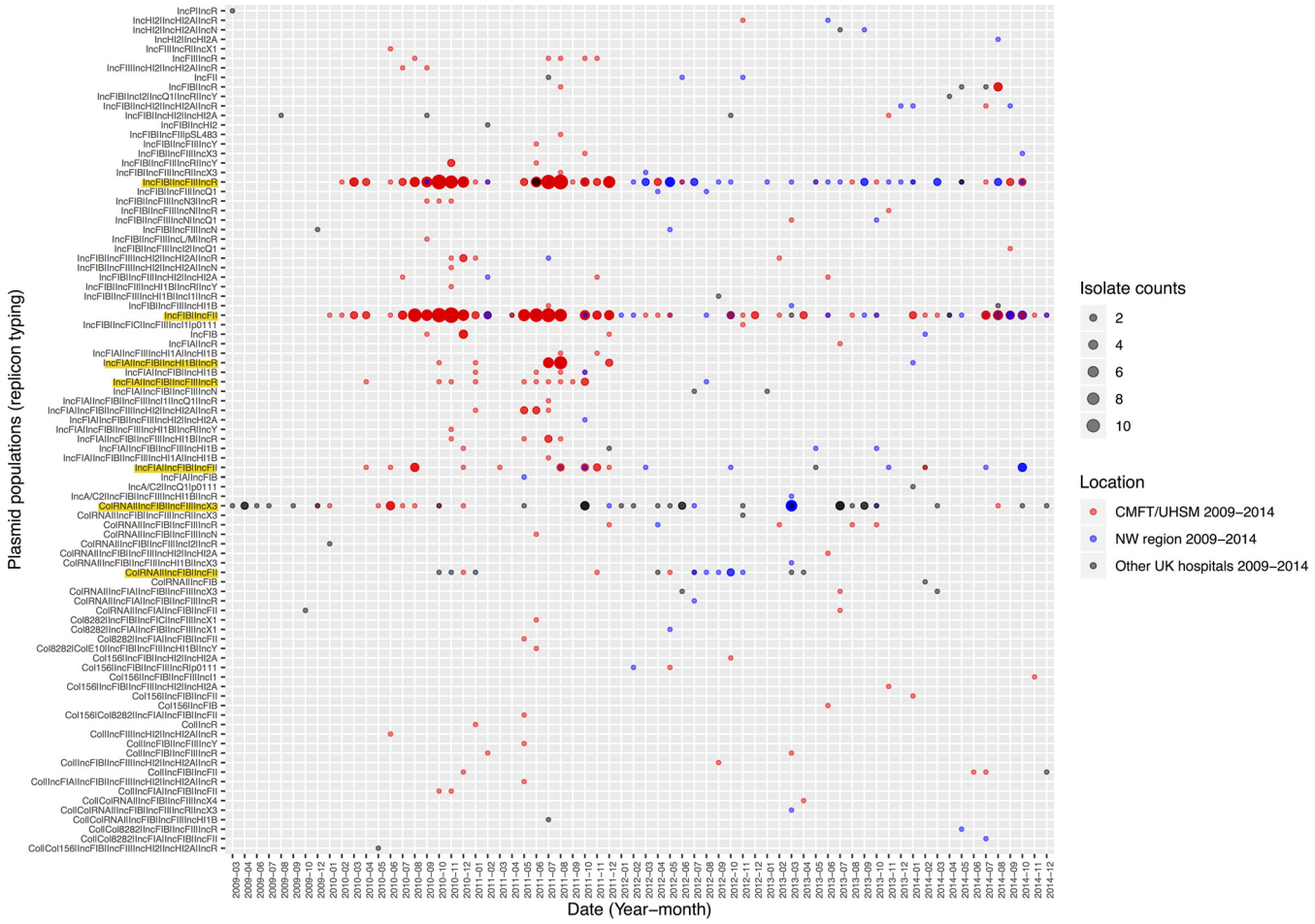

FIG 5 Incidence plot of plasmid populations identified in isolates (replicon typing) by year-month and geography. Dots are colored by location of isolate collection, as defined in Materials and Methods, and scaled by the number of isolates; the eight outpatient isolates were not plotted. The most predominant combinations are highlighted in yellow.

reference *bla*$_{KPC}$ plasmid database (see Materials and Methods), recognizing that any complete plasmid typing approach from short-read data is suboptimal; matches to one or more reference *bla*$_{KPC}$ plasmid sequences were identified in 554/604 (92%) isolates. Filtering the single match with the highest score at the predefined threshold of ≥0.80 left a subset of 428/554 (77%) for evaluation. These 428 isolates had matches to 12 *bla*$_{KPC}$ plasmid clusters (Fig. 6).

While the majority of isolates appeared to contain pKpQIL-like plasmids (323/604 [53%]), no significant matches to any reference plasmid were found in 162/604 (27%) isolates, suggesting that the genetic background supporting *bla*$_{KPC}$ in these isolates has diversified substantially and rapidly (Fig. 6). *bla*$_{KPC}$ plasmid cluster assignations were shared across a median (IQR) of 3 (1 to 6) STs, with pKpQIL-like plasmids being most widespread across species/STs (7 species, 75 STs) (Fig. 6) and clearly playing a major role in the dissemination of *bla*$_{KPC}$ in Manchester, North West England, and nationally over time (Fig. 7). Other plasmid types identified as top matches across the entire data set included those fully resolved by long-read sequencing performed in this study, some of which were seen in ≥5% of study isolates (e.g., pKPC-trace75 [a nontypeable replicon]) and in non-North West UK settings, likely reflecting recombination and the generation of new *bla*$_{KPC}$ plasmid variants in North West England and their subsequent dissemination (Fig. 7). Of note, among the 26 *E. cloacae* isolates sequenced, 12 (46%; five STs) had a match to the pKPC-272-like cluster, suggesting that *E. cloacae* is a particularly suitable host for this plasmid group (Fig. 6).

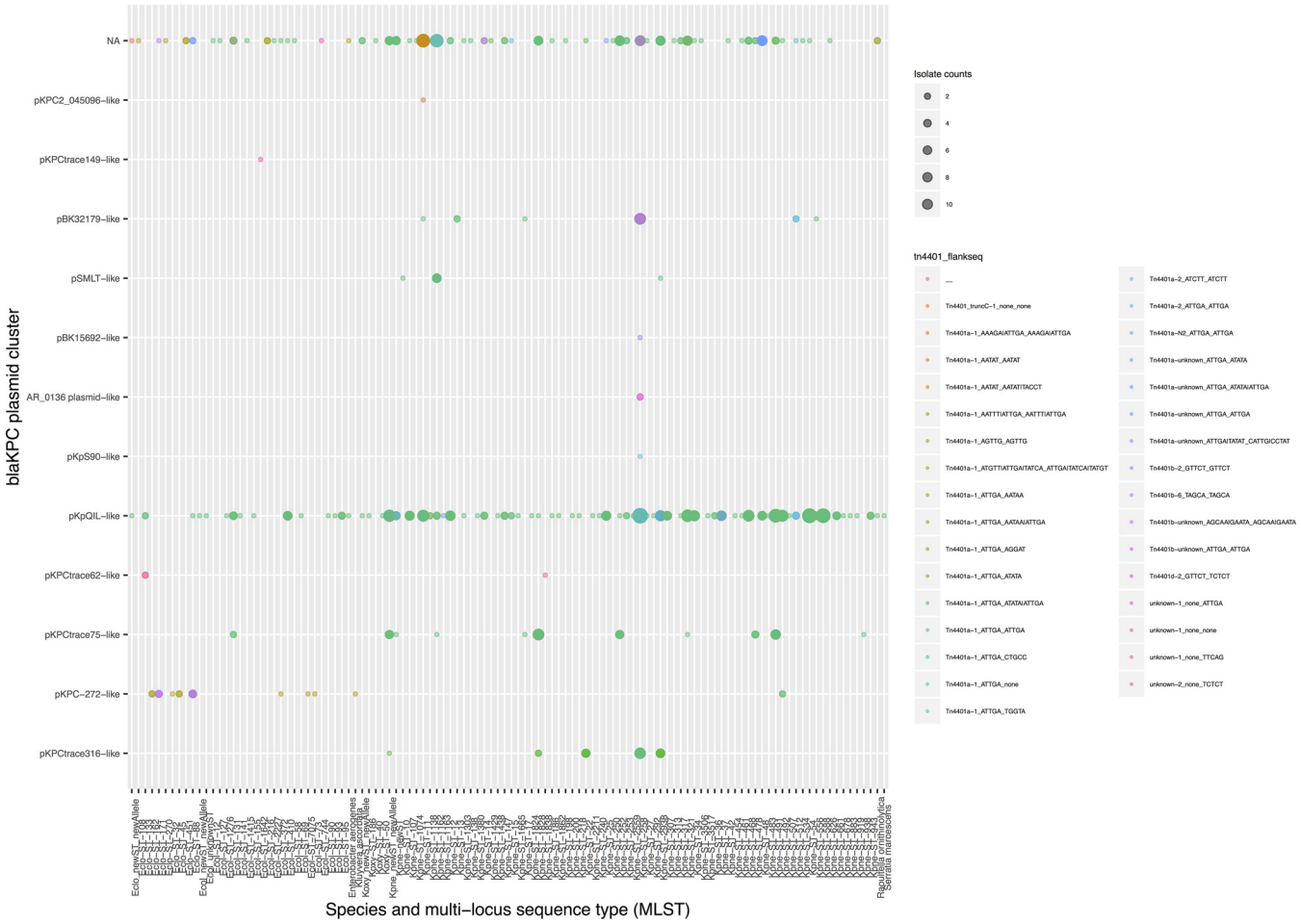

FIG 6 Distribution of *bla*$_{KPC}$ plasmid types by species-ST. Dots are colored by Tn*4401*/target site sequence type and scaled by the number of isolates. NA, not assigned.

Reconstruction of a subset of genomes using long-read sequencing data demonstrates a diverse population of plasmids with genetic rearrangement observed in both *bla*$_{KPC}$-positive and *bla*$_{KPC}$-negative cases. In addition to short-read data, to resolve genetic structures fully we obtained long-read PacBio data for 23 isolates, chosen to maximize the *bla*$_{KPC}$ plasmid diversity assayed and focusing on isolates collected from the two main Manchester hospitals (12 CMFT isolates and 5 UHSM isolates, plus 2 from other hospitals in North West England and 4 from other UK locations). These included the two earliest available *bla*$_{KPC}$ isolates from both CMFT and UHSM as well as isolates sharing the same species/ST but with different plasmid replicon combinations or from North West regional versus national locations, same-species isolates with different STs, and isolates of different species. One PacBio sequencing data set represented a clear isolate mixture (trace597 [UHSM] of *E. cloacae* ST133 and *K. pneumoniae* ST258), and for one isolate (trace457 [CMFT]) there were discrepancies between the short-read and long-read sequencing data sets, suggesting a laboratory error (*E. cloacae* ST45 long read, *E. coli* ST88 short read). These two assemblies were excluded, leaving 21 assemblies for further analysis (Data Set SD3).

Of the 153 contigs from these 21 assemblies, 30 were clearly chromosomal, 77 plasmid, 1 chromosomal with an integrated plasmid, and 45 with unclear provenance (i.e., possibly phage, plasmid, or chromosomal). Overall, 78/153 (51%) contigs were circularized, including 56/77 (73%) clear plasmid sequences. Thirty-one contigs (21 [68%] circularized) harbored *bla*$_{KPC}$, of which 21 were clearly plasmids (14/21 [67%] circularized), and one (trace552, *K. pneumoniae* ST11) had *bla*$_{KPC}$ integrated into the

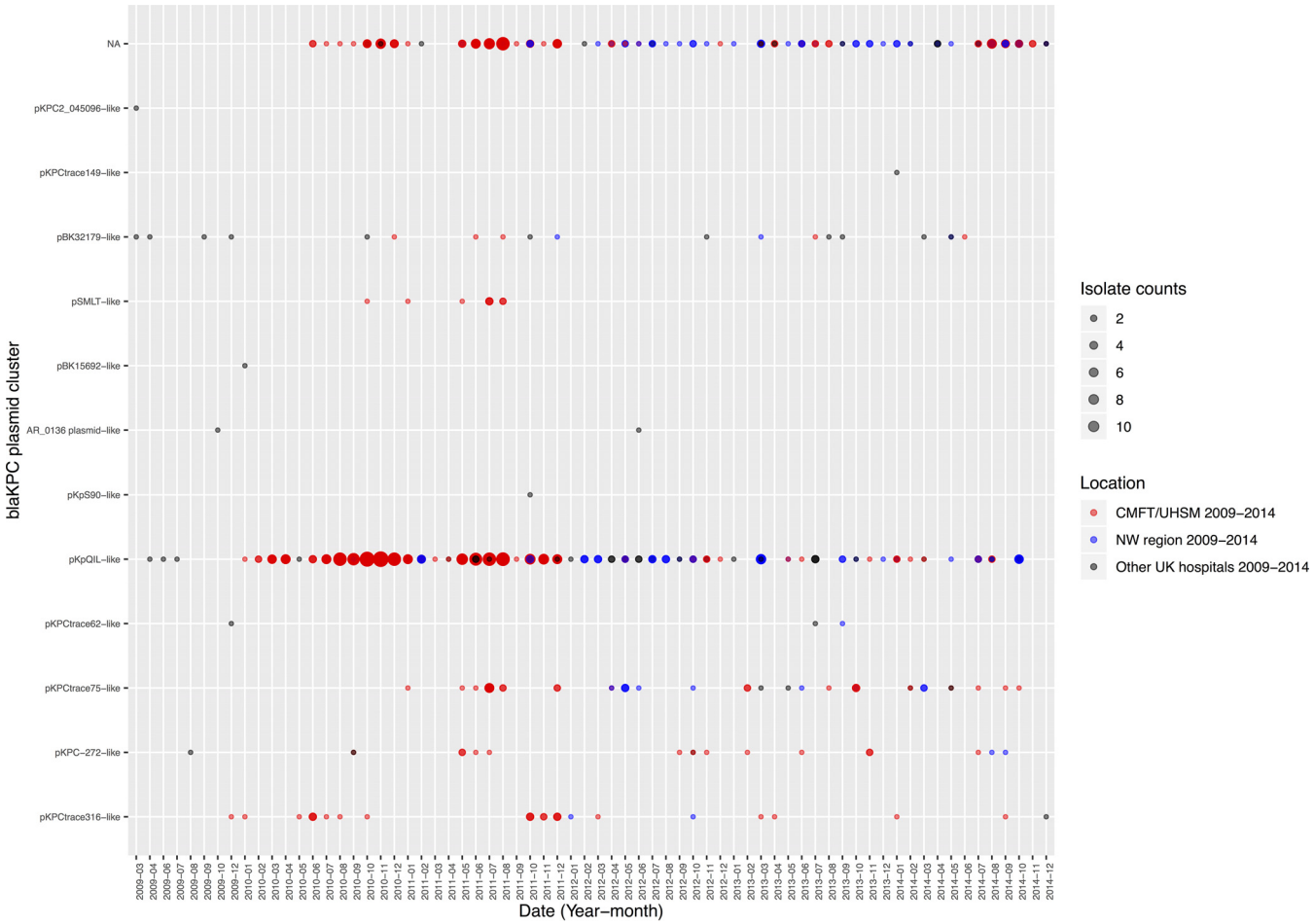

FIG 7 Incidence plot of bla_{KPC} plasmid types identified by year-month and geography. Dots are colored by location of isolate collection, as defined in Materials and Methods, and scaled by the number of isolates; the eight outpatient isolates were not plotted. NA, not assigned.

chromosome (not circularized). For nine other bla_{KPC}-positive contigs, we were not able to clearly ascertain whether these were chromosomal or plasmid (Data Set SD3).

We observed bla_{KPC} in multiple plasmid backgrounds (Fig. 8), including a majority of bla_{KPC} plasmids with multiple replicons (13/21 [60%] clear plasmid contigs, as represented in Fig. 8), particularly with IncFIB/IncFII and/or IncR, consistent with replicon patterns in the isolates overall (Fig. 5). For the IncFII group, for which we had 11 complete, reconstructed plasmid sequences with an IncFII(K)_CP000648-like replicon (plasmidFinder match; 3 bla_{KPC} negative [i.e., not represented in Fig. 8] and 8 bla_{KPC} positive), there was evidence of exchange and rearrangement of plasmid components between both bla_{KPC}-positive and bla_{KPC}-negative plasmids, as well as sharing between STs and species (Fig. 9). For example, between the bla_{KPC}-negative IncFII(K) plasmid isolated from a *Raoultella ornithinolytica* isolate and a bla_{KPC}-positive *K. pneumoniae* ST1828 isolate, >70 kb of sequence was ≥90% identical at the nucleotide level, as well as being largely syntenic [Fig. 9, annotation (i)]. A genetically related IncFII(K)/IncR plasmid was shared across species/lineages (*K. pneumoniae* [novel ST], *E. coli* [ST372], and *K. pneumoniae* [ST883]) and was similar to an IncFII(K)/IncFIB(pQIL) plasmid (found in *K. pneumoniae* ST1828 and *K. pneumoniae* ST588) [Fig. 9, annotation (ii)].

In addition to their plasticity, part of the success of these bla_{KPC} plasmids also may be attributable to the presence of toxin-antitoxin plasmid addiction systems (*ccdA-ccdB*, $n = 4$ bla_{KPC} plasmids; *higA*, $n = 6$; *vapB-vapC*, $n = 11$), antirestriction mechanisms (*klcA*, $n = 16$; previously shown to promote bla_{KPC} dissemination [20]), and heavy-metal resistance (*terB* [tellurite], $n = 3$; *ars* operon [arsenicals], $n = 3$; chromate resistance, $n = 1$; *cop* operon-*pcoC-pcoE* [copper], $n = 7$; *mer* operon [mercury], $n = 10$).

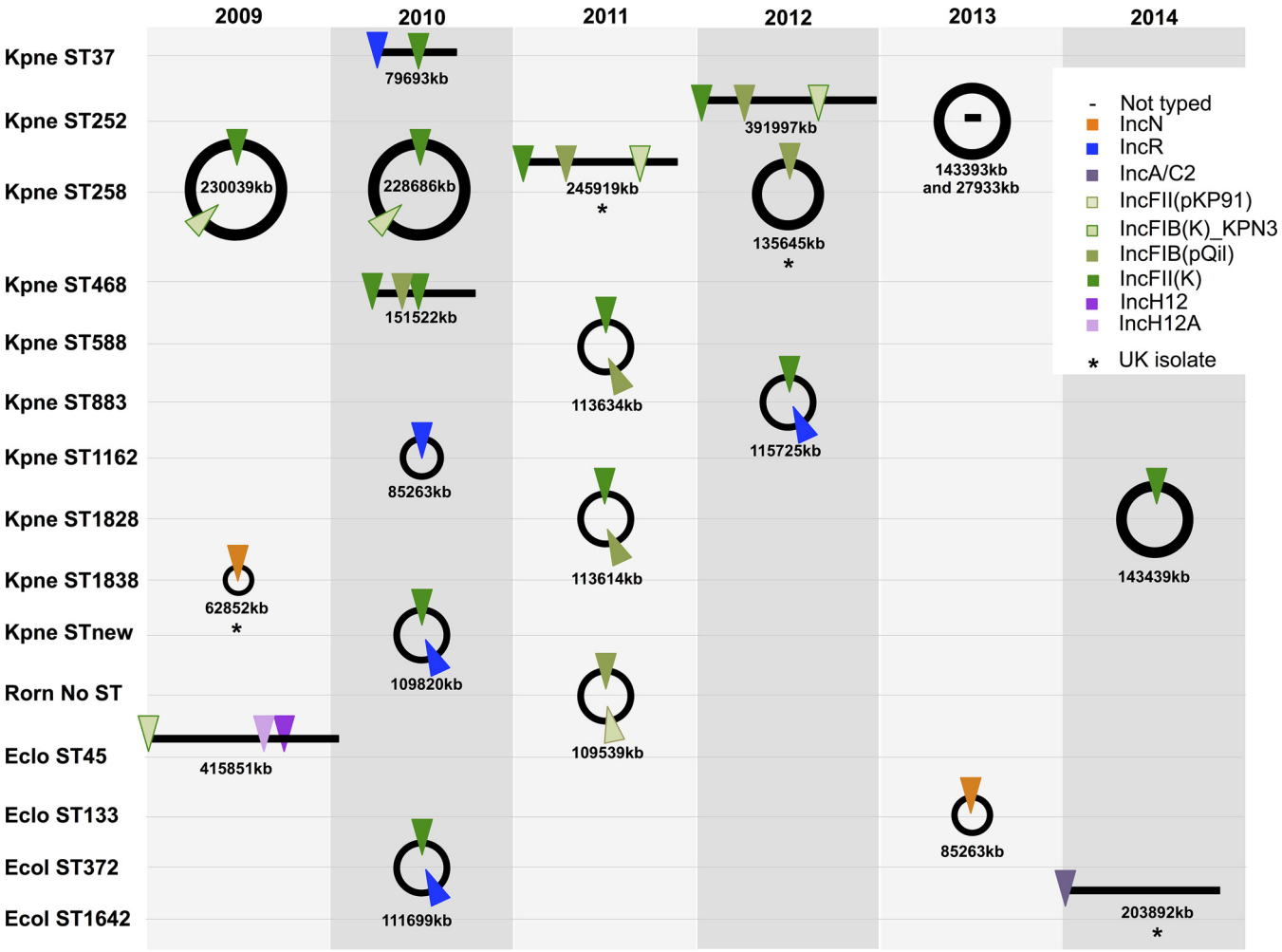

FIG 8 Schematic of bla_{KPC} plasmid types and sizes identified from long-read/short-read hybrid sequencing approach by species/ST and year of collection (only 21 contigs clearly designated plasmid are represented). Closed circles denote circularized contigs (i.e., complete plasmids); replicons are denoted by colored triangles in their approximate positions in the structure. Colored triangles denote replicon types assigned to each plasmid sequence (i.e., multiple colored triangles represent multireplicon plasmids). Plasmids from isolates from the wider UK collection (i.e., collected through the national reference laboratory) are denoted with an asterisk.

DISCUSSION

We present the largest WGS-based analysis of bla_{KPC}-positive isolates ($n = 604$) to our knowledge, focused on assessing genetic diversity around the carbapenemase gene itself, rather than limiting the analysis based on species type, and incorporating a sampling frame from UK regional and national collections over 5 years. bla_{KPC} remains one of the three most common carbapenemases observed in the United Kingdom, accounting for ~11% of cases referred to the AMRHAI Reference Unit in 2018 (OXA-48-like, 52%; NDM, 27%) (21) and presenting a significant challenge to hospitals in North West England, including Manchester, where it accounted for >97% of carbapenem resistance through 2015 (22).

Our study provides an interesting context in which to consider the findings of a recently published pan-European survey of carbapenem-nonsusceptible *K. pneumoniae* (the EuSCAPE study; 6 months, 2013 to 2014; 244 hospitals, 32 countries) (23). In EuSCAPE, 684 carbapenemase-producing *Klebsiella* species isolates were sequenced using Illumina technology, and, similar to our study, most cases were health care exposed (<2% from outpatients). EuSCAPE carbapenemase-producing isolates were also predominantly bla_{KPC} (~45%, $n = 311$ isolates) but mostly bla_{KPC-3} (232/311 [75%] versus 27/604 [5%] in our study) and ST258/ST512 (226/311 [73%] versus 107/525 [20%]

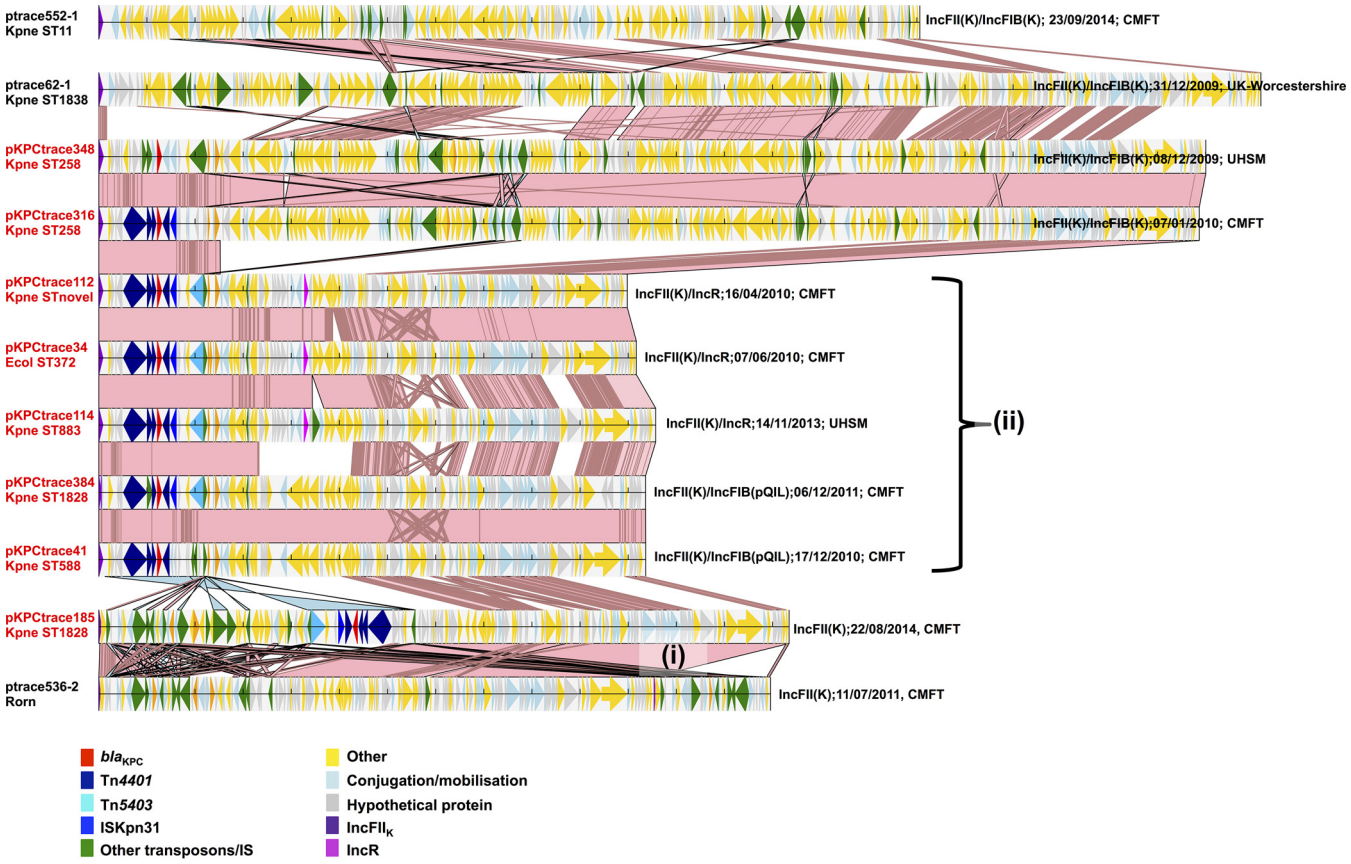

FIG 9 Alignments of complete plasmid sequences harboring an IncFII(K)_1_CP000648-like replicon, including bla_{KPC}-negative and bla_{KPC}-positive sequences. All sequences were reoriented to start at IncFII for the purposes of alignment visualization. Loci of interest have been colored and annotated as shown. Shading between sequences denotes regions of homology, with light pink shading denoting areas of ≥90% nucleotide identity, dark pink areas of ≥50% nucleotide identity, and light blue areas of ≥90% nucleotide identity in reverse orientation. The order of sequences is adjusted to highlight genetic overlap between sequences but not to imply any specific direct exchange events. Annotations (i) and (ii) denote specific features highlighted in the main text.

of *K. pneumoniae* overall in our study). Based on identifying genetic nearest neighbors in their data, the EuSCAPE team found 51% of bla_{KPC} *K. pneumoniae* isolates were most closely related to another isolate from the same hospital. The authors concluded that there was strong geographic structuring of strains, and that the expansion of a few clonal lineages was predominantly responsible for the spread of carbapenemases in *K. pneumoniae* in Europe, with onward nosocomial transmission. Like bla_{KPC-3} in EuSCAPE, bla_{KPC-2} also has been linked with the clonal expansion of ST258 in Australia (24), where 48% of 176 *K. pneumoniae* isolates sequenced were bla_{KPC-2}-containing ST258.

However, instead of clonal expansion, as found in EuSCAPE, in our study we found rapid dissemination of mobile backgrounds supporting bla_{KPC-2}, similar to observations from sequencing of other polyclonal bla_{KPC} scenarios reported elsewhere, including the United States (5, 25). Tn*4401a*, associated with high levels of bla_{KPC} expression (26), previously was seen predominantly in *K. pneumoniae* and in isolates from the United States, Israel, and Italy, and, similarly, most commonly with an ATTGA-ATTGA TSD (9). Thus, our findings are consistent with the importation of the predominant bla_{KPC-2}-Tn*4401a*-ATTGA-ATTGA motif into CMFT/North West England and subsequent horizontal spread. Notably, as in EuSCAPE, 46/72 (64%) singleton isolates we sampled from UK hospitals were also CG258, but our much more detailed sampling across species reflected a very different local/regional molecular epidemiology. Although the EuSCAPE study is large and impressive, its breadth may have been limiting in understanding regional diversity; for example, the bla_{KPC} *K. pneumoniae* subset from the United Kingdom that was analyzed in EuSCAPE consisted of 11 isolates submitted from six

centers (https://microreact.org/project/EuSCAPE_UK). The focus also was more on analyzing species-specific clonal relationships, with no analysis of other species or MGEs.

Although in our study diversification occurred at all genetic levels (Tn*4401* plus TSS, plasmids, plasmid populations, strains, and species), there was more limited variation observed within the Tn*4401* transposon and its flanking regions, and the spread of bla_{KPC} appears to have been supported by highly plastic modular exchange of larger genetic segments within a distinct plasmid population, particularly IncFIB/IncFII (found in 580 and 545 of the 604 isolates, respectively) and IncR replicons (252/604 isolates). A previous study, in which 11 transformed bla_{KPC} plasmids from the United Kingdom (2008 to 2010) were sequenced (Roche 454/assembly, PCR plus sequencing-based gap closure), identified a UK variant of the pKpQIL plasmid, designated pKpQIL-UK ($IncFII_{K2}$ by plasmid MLST), that was highly similar to pKpQIL (maximum SNV diversity of 32) and several other $IncFII_{K2}$ pKpQIL-like plasmids but with novel segmental genetic rearrangements (gains/losses; pKpQIL-D1 and pKpQIL-D2) (14). Our data also support the importance of $IncFII_{K2}$-like plasmids in bla_{KPC} dissemination and that other $IncFII_K$-like plasmids (e.g., $IncFII_{K1, -K4, -K7, -K15}$) and replicons (IncFIB and IncR) have been a significant feature. In addition to their plasticity, the plasmids identified frequently harbored AMR genes other than bla_{KPC}, which might offer a selective advantage, alongside heavy-metal resistance genes and plasmid toxin-antitoxin addiction systems. The plasticity and association of $IncFII_K$ plasmids with resistance genes and IncFIB replicons has been supported by findings of a recent analysis of $IncFII_K$ plasmids (27).

The problem of accurately classifying plasmid populations from short-read data was exemplified in this analysis and highlighted by our smaller long-read/short-read hybrid assembly-based analysis, which demonstrated significant diversity within structures assigned as similar by short-read-based typing approaches. With this caveat, it was interesting that even with relatively relaxed thresholds, 29% of isolates did not have a match to our reference bla_{KPC} plasmid database (based on clustering of all publicly available reference sequences, as in Materials and Methods), consistent with rapid diversification in the plasmid backgrounds supporting Tn*4401*/bla_{KPC-2} in this setting.

Our findings demonstrated that it is also important to consider plasmids without the resistance gene of interest in a population, as these may be relevant to a wider understanding of the transmission and evolution of smaller mobile genetic elements harboring resistance genes (Fig. 9). This was also shown to be relevant in a previous analysis of a large KPC-*E. coli* outbreak in the same setting in 2015 to 2016, in which a circulating bla_{KPC}-negative plasmid, pCAD3 (IncFIB/FII), acquired Tn*4401* from an IncHI2/HI2A bla_{KPC}-positive plasmid and went on to dominate within a clonal *E. coli* lineage (22). Most studies in general, however, tend to focus on analyzing AMR plasmids of interest. Fortunately, long-read sequencing is becoming increasingly low cost and high throughput, and hybrid assembly is able to reconstruct plasmid sequences in *Enterobacterales* (28, 29). New developments in large-scale comparative genomics of complete genomes, including plasmid structures, are essential for future large-scale analysis of AMR gene transmission.

There are several limitations to our study. We were only able to long-read sequence a small number of isolates, and the reconstructed genomes generated using long-read PacBio data remained incomplete (49% of all contigs uncircularized). Improvements in long-read technology and assembly approaches will likely overcome this (28). Our short-read and long-read data sets were generated from the same frozen stocks of isolates but from separate subcultures (because we used the short-read data to inform selection for long-read sequencing); ideally they would have been generated from the same DNA extract. PacBio sequencing library preparation incorporates size selection, and this may have led to short plasmid sequences (<15 kb) being lost. Our interpretation of the evolution of backgrounds supporting bla_{KPC} was limited by the diversity present and the inability to capture sequential evolutionary events, even with this large study. We restricted our WGS reference-based plasmid typing to analyzing top matches to our bla_{KPC} plasmid database because of the uncertainty in using short-read data for plasmid typing (19); therefore, we may have underestimated the diversity of bla_{KPC}

plasmids present and missed cases with >1 bla_{KPC} plasmid. Lastly, very limited epidemiological data linked to the isolates were available, meaning that we were unable to ascertain any epidemiological drivers that might be contributing to the enormous heterogeneity of bla_{KPC} transmission over apparently short time frames; the latter finding also precluded the useful application of standard phylogenetic approaches based on identifying variants core to and within species. In addition, the collection of isolates by PHE as part of regional and national surveillance was dictated by referral patterns of isolates from the hospitals surveyed, and we do not have any denominator information on cultures (either bla_{KPC} positive or bla_{KPC} negative) to corroborate details on the robustness of this referral process or to determine what proportion of all UK bla_{KPC}-positive *Enterobacterales* over the relevant time frame we have sequenced.

In conclusion, our large analysis highlights the difficulty and complexity of bacterial transmission networks once important AMR genes have "escaped" the genetic confines of particular mobile genetic elements and bacterial species/lineages, with important implications for surveillance. These include the need to consider multiple bacterial species and plasmids as potential hosts of bla_{KPC} and invest resources in sequencing approaches to adequately reconstruct genetic structures and avoid misinterpreting the molecular epidemiology. It also demonstrates that regional differences in AMR gene epidemiology may be quite marked, which may affect the generalizability of control methods. Finally, it is important to consider the wider genetic background of host strains and plasmids in understanding the evolution and dissemination of important AMR genes, as AMR gene transfer between plasmid backgrounds within bacteria may occur over short timescales, and the interaction of several plasmids (i.e., not just those harboring the AMR gene of interest at any given time) in a population may be highly relevant to the persistence and dissemination of the AMR gene itself.

MATERIALS AND METHODS

Study isolates and setting. We sequenced archived carbapenem-resistant *Enterobacterales* isolates from two large teaching hospitals in Manchester (formerly known as CMFT and UHSM), aiming to include all inpatient isolates archived following the observed introduction of bla_{KPC}-positive *Enterobacterales* (KPC-E) in this hospital system in 2009 to 2011. We also sequenced a subset of (KPC-E) isolates archived and sequenced as part of regional and national surveillance of carbapenemases undertaken by Public Health England (PHE; 2009 to 2014). The PHE set included (i) a further random set of isolates referred from CMFT/UHSM from 2012 to 2014; (ii) up to the first 25 consecutive KPC-E isolates from any hospital in North West England (2009 to 2014) and referred to the PHE reference laboratory (2009 to 2014); (iii) the first KPC-E isolate from any other hospital in the United Kingdom and Ireland referred to PHE (2009 to 2014); and (iv) any KPC-E isolates from outpatient/primary care settings in the United Kingdom referred to PHE (2009 to 2014). For the UHSM/CMFT isolate subset, we were able to determine sampling and clinical sample culture-positivity denominators from an anonymized database of linked electronic bacteriology and patient administration records going back to 2009 (22).

Ethical approval was not required as only bacterial isolates were sequenced, and their collection was part of infection control investigation and management.

DNA extraction and sequencing. For short-read Illumina sequencing (150-bp PE reads; HiSeq 2500), DNA was extracted using QuickGene (Fujifilm, Japan), with an additional mechanical lysis step following chemical lysis (FastPrep; MP Biomedicals, USA). Sequencing libraries were constructed using the NEBNext Ultra DNA sample prep master mix kit (NEB), with minor modifications, and a custom automated protocol on a Biomek FX (Beckman). The ligation of adapters was performed using Illumina multiplex adapters, and ligated libraries were size selected using AMPure magnetic beads (Agencourt). Each library was PCR enriched with custom primers (index primer plus dual-index PCR primer) (30). Enrichment and adapter extension of each preparation was obtained using 9 μl from a size-selected library in a 50-μl PCR. Reaction mixtures then were purified with AMPure beads (Agencourt/Beckman) on a Biomek NXp after 10 cycles of amplification (per Illumina recommendations). Final size distributions of libraries were determined using a Tapestation 1DK system (Agilent/Lab901) and quantified by Qubit fluorometry (Thermofisher).

For long-read sequencing (PacBio [$n = 28$]), DNA was extracted using the Qiagen genomic tip 100/G kit (Qiagen, Netherlands). DNA extracts were initially sheared to an average length of 15 kb using g-tubes, as specified by the manufacturer (Covaris). Sheared DNA was used in SMRTbell library preparation, as recommended by the manufacturer. The quantity and quality of the SMRTbell libraries were evaluated using the high-sensitivity dsDNA kit and Qubit fluorimeter (Thermo Fisher Scientific) and DNA 12000 kit on the 2100 Bioanalyzer (Agilent). To obtain the longest possible SMRTbell libraries for sequencing (as recommended by the manufacturer), a further size selection step was performed using the PippinHT pulsed-field gel electrophoresis system (Sage Science), enriching for SMRTbell libraries of >15 kb for loading onto the instrument. Sequencing primer and P6 polymerase were annealed and bound to the

SMRTbell libraries, and each library was sequenced using a single SMRT cell on the PacBio RSII sequencing system.

Sequence data processing and assembly. We used Kraken (31) to assign species to sequenced isolates from short-read (Illumina) data. SPAdes v3.6 (32) was used to *de novo* assemble short reads (default options; subsequent removal of contigs shorter than 500 bp and assembly coverage of <2×). Isolates with short-read sequence assemblies of >6.5 Mb were excluded to ensure that potentially mixed sequences were not included in the analyses. MLST was derived *in silico* from short-read assemblies by *de novo* blasting these against publicly available MLST databases for *E. coli* (http://mlst.warwick.ac.uk/mlst/dbs/Ecoli), *K. pneumoniae*, *E. cloacae*, and *K. oxytoca* (https://pubmlst.org/). Isolates with mixed MLST outputs were excluded. Antimicrobial resistance (AMR) genes, plasmid replicon (Inc) types, and insertion sequences (IS) were identified from short-read data using resistType (https://github.com/hangphan/resistType_docker; curated AMR gene database as in reference 33, plasmid replicon reference sequences from PlasmidFinder [34], and ISs from the ISFinder platform [35]; ≥80% identity was used as a threshold). bla_{KPC} copy number (per bacterial genome) for each isolate was estimated from short-read assemblies by dividing coverage of the contig containing bla_{KPC} by the average coverage for the assembly (weighted by contig length). Confirmation that this was a biologically meaningful estimate was obtained by estimating the association between copy number estimates and meropenem MICs from routine clinical laboratory antimicrobial susceptibility testing using interval regression [outcome $\log_2$(MIC), left and right censored to reflect the actual MIC in the extreme categories being unobserved but only within a range] as the dependent variable and KPC copy number as the independent variable. bla_{KPC} copy number estimates also were validated by mapping Illumina reads to the reconstructed genomes derived from the PacBio data (using bwa-mem [bwa-0.7.12-r1034] [36] only where both chromosome and bla_{KPC} plasmid structures were deemed complete) and then by calculating mean coverage for the bla_{KPC} gene versus the chromosomal contig and comparing this with the bla_{KPC} copy number estimate derived from the short-read assembly described above. Plasmid MLST for common family types identified in short-read data was confirmed by 100% sequence matches to reference alleles for families catalogued on the plasmidMLST website (https://pubmlst.org/plasmid/) (IncA/C, IncHI1/2, and IncN).

For the smaller subset of isolates on which long-read (i.e., PacBio) sequencing was performed, long-read data were assembled using the HGAP pipeline (37) and polished with the corresponding Illumina data sets using Pilon (version 1.18; default parameters) (38). Chromosomal sequences and plasmid sequences then were manually curated where possible to create complete, closed, circular structures by using BLASTn to identify overlaps at the end of assembled contigs. Those with overlapping ends larger than 1,000 bp with sequence identity of >99% were considered circularized/complete and trimmed appropriately for resolution. Complete sequences were annotated using PROKKA (version 1.11) (39); annotations were used to determine genes known to encode toxin-antitoxin systems, heavy-metal resistance, and antirestriction mechanisms. Plasmid MLST was confirmed for these assemblies as described above.

Tn*4401* typing. Tn*4401* typing was performed using TETyper (9), using the Tn*4401*, single-nucleotide polymorphism, and structural profile reference files included with the package (https://github.com/aesheppard/TETyper, version 1.1), as well as a flanking length of 5 bp, representative of the known target site signature sequence indicative of Tn*4401* transposition (40).

Plasmid database for bla_{KPC} plasmid typing. A reference bla_{KPC} plasmid sequence database composed of bla_{KPC}-harboring contigs/plasmids from long-read sequencing of isolates in this study and all complete bla_{KPC} plasmids from references 41–43 (August 2018) was used for bla_{KPC} plasmid typing in this study. To construct this database, all 279/6,018 evaluable plasmid sequences carrying bla_{KPC} first were compared using *dnadiff* (44) to obtain the pairwise similarity between any two plasmid sequences, p_i and p_j. The similarity was defined as a function of their lengths, l_i and l_j, and the aligned bases, l_{ij} and l_{ji}, as reported by the equation

$$(p_i, p_j) = \frac{1}{2}\left(\frac{l_{ij}}{l_i} + \frac{l_{ji}}{l_j}\right) \times \min\left(\frac{l_i}{l_j}, \frac{l_j}{l_i}\right)$$

The score was designed to penalize differences in lengths of the compared sequences, i.e., to make sequences of different lengths proportionately more different. The resulting similarity matrix was used to perform clustering of plasmid sequences using the affinity propagation clustering technique, suitable for graph clustering problems with sparse similarity matrix and uneven cluster size and cluster number (45), and it resulted in 34 clusters of 1 to 43 plasmids per cluster (see Data Set SD4 in the supplemental material). The largest cluster was the set of pKpQIL-like plasmids comprising 43 related sequences. Representative sequences of each bla_{KPC} plasmid cluster in this network were chosen randomly to generate a set (*KPC-pDB*) of plasmids ranging from 7,995 bp (GenBank accession no. NC_022345.1; plasmid pAP-2) to 447,095 bp (NZ_CP029436.1; plasmid pKPC_CAV2013) in the final database used for bla_{KPC} plasmid typing in this study.

Subsequently, bla_{KPC} plasmid typing for each study isolate sequence was performed as follows. (i) Assembled sequences for each isolate were subjected to BLAST (BLASTn) searches against *KPC-pDB*. (ii) Any >1-kb contig with >90% nucleotide identity and >80% total coverage match to sequences in *KPC-pDB* was retained. (iii) For any sequence p_i in *KPC-pDB*, a score, s_i, was calculated by dividing the total matched bases of all contigs matched to p_i by p_i's length. (iv) An isolate was assumed to plausibly carry p_i if s_i was ≥0.80. An isolate could have several bla_{KPC} plasmid matches; we restricted to the top match for each isolate in our analyses.

Phylogenetic reconstruction for isolates from commonly represented STs. To ensure that for the most common STs (*K. pneumoniae* STs 258, 11, and 491) we were essentially not characterizing a single, highly clonal set of sequences, recombination-corrected phylogenies were reconstructed for each ST

following mapping to the *K. pneumoniae* reference MGH78578 (GenBank accession no. CP000647.1), as previously described (46). In brief, following read mapping, variant calling, and generation of consensus fasta sequences of variants called, IQtree followed by ClonalFrameML were run using a python wrapper, available at https://github.com/davideyre/runListCompare (accessed January 2020).

Statistical analysis and data visualization. Statistical analysis (Kruskal-Wallis tests, interval regression, and Pearson's correlation) was carried out in Stata 14.2. Plots for Fig. 1 and 7 and supplemental figures were generated using the ggplot2 and ape packages in R (version 1.1.463). Figure 9 was generated using the GenomeDiagram package (47) in Biopython (48).

Sequencing data availability. Sequencing data have been deposited in the NCBI database (BioProject accession no. PRJNA564424). PacBio/Pilon assemblies are available at https://doi.org/10.6084/m9.figshare.11777631.v1. Typing results and metadata for each isolate are available in Data Set SD1.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.6 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.1 MB.

SUPPLEMENTAL FILE 3, XLSX file, 0.04 MB.

SUPPLEMENTAL FILE 4, XLSX file, 0.1 MB.

SUPPLEMENTAL FILE 5, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

We are grateful to and acknowledge the sharing of isolates by microbiology and clinical teams from contributing UK hospitals and from Martin Cormican, the National Reference Laboratory in Galway, Ireland, and the contributing laboratories in Ireland. We are also grateful to the microbiology laboratory staff and infection control teams at Manchester University NHS Foundation Trust (formerly CMFT and UHSM); the staff of the Manchester Medical Microbiology Partnership; and the research laboratory, informatics, and project management teams working as part of the Modernizing Medical Microbiology Consortium, Oxford.

Contemporaneous investigation by CMFT, UHSM, and PHE was undertaken as part of routine activity. The retrospective investigation was funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Healthcare-Associated Infections and Antimicrobial Resistance at Oxford University in partnership with Public Health England (PHE) (grant HPRU-2012-10041) and supported by the NIHR Biomedical Research Centre, Oxford. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, the Department of Health, or Public Health England.

N.S. is funded by a PHE/University of Oxford Academic Clinical Lectureship. T.E.A.P., D.W.C., and A.S.W. are NIHR Senior Investigators.

The Transmission of Carbapenemase-Producing *Enterobacteriaceae* (TRACE) study investigators are listed alphabetically and include several of the authors also listed by name in the main author list: Zoie Aiken, Oluwafemi Akinremi, Aiysha Ali, Julie Cawthorne, Paul Cleary, Derrick W. Crook, Valerie Decraene, Andrew Dodgson, Michel Doumith, Matthew J. Ellington, Ryan George, John Grimshaw, Malcolm Guiver, Robert Hill, Katie L. Hopkins, Rachel Jones, Cheryl Lenney, Amy J. Mathers, Ashley McEwan, Ginny Moore, Mark Neilson, Sarah Neilson, Timothy E. A. Peto, Hang T. T. Phan, Mark Regan, Anna C. Seale, Nicole Stoesser, Jay Turner-Gardner, Vicky Watts, A. Sarah Walker, Jimmy Walker, David Wyllie, William Welfare, and Neil Woodford.

None of the authors has any conflicts of interest to declare.

REFERENCES

1. Mathers AJ, Stoesser N, Sheppard AE, Pankhurst L, Giess A, Yeh AJ, Didelot X, Turner SD, Sebra R, Kasarskis A, Peto T, Crook D, Sifri CD. 2015. Klebsiella pneumoniae carbapenemase (KPC)-producing K. pneumoniae at a single institution: insights into endemicity from whole-genome sequencing. Antimicrob Agents Chemother 59:1656–1668. https://doi.org/10.1128/AAC.04292-14.
2. Cerqueira GC, Earl AM, Ernst CM, Grad YH, Dekker JP, Feldgarden M, Chapman SB, Reis-Cunha JL, Shea TP, Young S, Zeng Q, Delaney ML, Kim D, Peterson EM, O'Brien TF, Ferraro MJ, Hooper DC, Huang SS, Kirby JE, Onderdonk AB, Birren BW, Hung DT, Cosimi LA, Wortman JR, Murphy CI, Hanage WP. 2017. Multi-institute analysis of carbapenem resistance reveals remarkable diversity, unexplained mechanisms, and limited clonal outbreaks. Proc Natl Acad Sci U S A 114:1135–1140. https://doi.org/10.1073/pnas.1616248114.
3. Ruiz-Garbajosa P, Curiao T, Tato M, Gijón D, Pintado V, Valverde A, Baquero F, Morosini MI, Coque TM, Cantón R. 2013. Multiclonal dispersal

of KPC genes following the emergence of non-ST258 KPC-producing Klebsiella pneumoniae clones in Madrid, Spain. J Antimicrob Chemother 68:2487–2492. https://doi.org/10.1093/jac/dkt237.
4. Martin J, Phan H, Findlay J, Stoesser N, Pankhurst L, Navickaite I, De Maio N, Eyre D, Toogood G, Orsi N, Kirby A, Young N, Turton J, Hill R, Hopkins K, Woodford N, Peto T, Walker A, Crook D, Wilcox M. 2017. Covert dissemination of carbapenemase-producing Klebsiella pneumoniae (KPC) in a successfully controlled outbreak: long and short-read whole-genome sequencing demonstrate multiple genetic modes of transmission. J Antimicrob Chemother 72:3025–3034. https://doi.org/10.1093/jac/dkx264.
5. Sheppard AE, Stoesser N, Wilson DJ, Sebra R, Kasarskis A, Anson LW, Giess A, Pankhurst LJ, Vaughan A, Grim CJ, Cox HL, Yeh AJ, Modernising Medical Microbiology Informatics Group, Sifri CD, Walker AS, Peto TE, Crook DW, Mathers AJ. 2016. Nested Russian doll-like genetic mobility drives rapid dissemination of the carbapenem resistance gene blaKPC. Antimicrob Agents Chemother 60:3767–3778. https://doi.org/10.1128/AAC.00464-16.
6. Quainoo S, Coolen JPM, van Hijum S, Huynen MA, Melchers WJG, van Schaik W, Wertheim H. 2017. Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. Clin Microbiol Rev 30:1015–1063. https://doi.org/10.1128/CMR.00016-17.
7. Walker TM, Monk P, Smith EG, Peto TE. 2013. Contact investigations for outbreaks of Mycobacterium tuberculosis: advances through whole genome sequencing. Clin Microbiol Infect 19:796–802. https://doi.org/10.1111/1469-0691.12183.
8. Munoz-Price LS, Poirel L, Bonomo RA, Schwaber MJ, Daikos GL, Cormican M, Cornaglia G, Garau J, Gniadkowski M, Hayden MK, Kumarasamy K, Livermore DM, Maya JJ, Nordmann P, Patel JB, Paterson DL, Pitout J, Villegas MV, Wang H, Woodford N, Quinn JP. 2013. Clinical epidemiology of the global expansion of Klebsiella pneumoniae carbapenemases. Lancet Infect Dis 13:785–796. https://doi.org/10.1016/S1473-3099(13)70190-7.
9. Sheppard AE, Stoesser N, German-Mesner I, Vegesana K, Sarah Walker A, Crook DW, Mathers AJ. 2018. TETyper: a bioinformatic pipeline for classifying variation and genetic contexts of transposable elements from short-read whole-genome sequencing data. Microb Genom 4:e000232. https://doi.org/10.1099/mgen.0.000232.
10. Palepou M, Hope R, Colman M, Glover J, Kaufmann M, Lafong C, Reynolds R, Livermore DM. 2005. Novel class A carbapenemase, KPC-4, in an Enterobacter isolate from Scotland, abstr 1134_01_20. Prog Abstr 15th Eur Cong Clin Microbiol Infect Dis.
11. Public Health England. 2011. Carbapenemase-producing Enterobacteriaceae: laboratory confirmed cases, 2003 to 2013. https://www.gov.uk/government/publications/carbapenemase-producing-enterobacteriaceae-laboratory-confirmed-cases/carbapenemase-producing-enterobacteriaceae-laboratory-confirmed-cases-2003-to-2013. Accessed 9 February 2016.
12. Donker T, Henderson KL, Hopkins KL, Dodgson AR, Thomas S, Crook DW, Peto TEA, Johnson AP, Woodford N, Walker AS, Robotham JV. 2017. The relative importance of large problems far away versus small problems closer to home: insights into limiting the spread of antimicrobial resistance in England. BMC Med 15:86. https://doi.org/10.1186/s12916-017-0844-2.
13. Findlay J, Hopkins KL, Doumith M, Meunier D, Wiuff C, Hill R, Pike R, Loy R, Mustafa N, Livermore DM, Woodford N. 2016. KPC enzymes in the UK: an analysis of the first 160 cases outside the North-West region. J Antimicrob Chemother 71:1199–1206. https://doi.org/10.1093/jac/dkv476.
14. Doumith M, Findlay J, Hirani H, Hopkins KL, Livermore DM, Dodgson A, Woodford N. 2017. Major role of pKpQIL-like plasmids in the early dissemination of KPC-type carbapenemases in the UK. J Antimicrob Chemother 72:2241–2248. https://doi.org/10.1093/jac/dkx141.
15. Qi Y, Wei Z, Ji S, Du X, Shen P, Yu Y. 2011. ST11, the dominant clone of KPC-producing Klebsiella pneumoniae in China. J Antimicrob Chemother 66:307–312. https://doi.org/10.1093/jac/dkq431.
16. Cuzon G, Naas T, Nordmann P. 2011. Functional characterization of Tn4401, a Tn3-based transposon involved in blaKPC gene mobilization. Antimicrob Agents Chemother 55:5370–5373. https://doi.org/10.1128/AAC.05202-11.
17. Cuzon G, Naas T, Truong H, Villegas MV, Wisell KT, Carmeli Y, Gales AC, Venezia SN, Quinn JP, Nordmann P. 2010. Worldwide diversity of Klebsiella pneumoniae that produce beta-lactamase blaKPC-2 gene. Emerg Infect Dis 16:1349–1356. https://doi.org/10.3201/eid1609.091389.
18. He S, Hickman AB, Varani AM, Siguier P, Chandler M, Dekker JP, Dyda F. 2015. Insertion sequence IS26 reorganizes plasmids in clinically isolated multidrug-resistant bacteria by replicative transposition. mBio 6:e00762-15. https://doi.org/10.1128/mBio.00762-15.
19. Arredondo-Alonso S, Willems RJ, van Schaik W, Schürch AC. 2017. On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. Microb Genom 3:e000128. https://doi.org/10.1099/mgen.0.000128.
20. Liang W, Xie Y, Xiong W, Tang Y, Li G, Jiang X, Lu Y. 2017. Anti-restriction protein, KlcAHS, promotes dissemination of carbapenem resistance. Front Cell Infect Microbiol 7:150. https://doi.org/10.3389/fcimb.2017.00150.
21. Public Health England. 2018. English Surveillance Programme for Antimicrobial Utilisation and Resistance (ESPAUR), report 2018. Public Health England, London, United Kingdom.
22. Decraene V, Phan HTT, George R, Wyllie DH, Akinremi O, Aiken Z, Cleary P, Dodgson A, Pankhurst L, Crook DW, Lenney C, Walker AS, Woodford N, Sebra R, Fath-Ordoubadi F, Mathers AJ, Seale AC, Guiver M, McEwan A, Watts V, Welfare W, Stoesser N, Cawthorne J, Aiken Z, Akinremi O, Cawthorne J, Cleary P, Crook DW, Decraene V, Dodgson A, Doumith M, Ellington M, Eyre DW, George R, Guiver M, Hill R, Hopkins K, Jones R, Lenney C, Mathers AJ, McEwan A, Moore G, Neilson S, Peto TEA, Phan HTT, Regan M, Seale AC, Stoesser N, Turner-Gardner J, Watts V, Walker J, Walker AS, Wyllie DH, Welfare W, Woodford N, TRACE Investigators' Group. 2018. A large, refractory nosocomial outbreak of Klebsiella pneumoniae carbapenemase-producing Escherichia coli demonstrates carbapenemase gene outbreaks involving sink sites require novel approaches to infection control. Antimicrob Agents Chemother 62:e01689-18. https://doi.org/10.1128/AAC.01689-18.
23. David S, Reuter S, Harris SR, Glasner C, Feltwell T, Argimon S, Abudahab K, Goater R, Giani T, Errico G, Aspbury M, Sjunnebo S, Eu SWG, EuSCAPE Working Group, ESGEM Study Group, Feil EJ, Rossolini GM, Aanensen DM, Grundmann H. 2019. Epidemic of carbapenem-resistant Klebsiella pneumoniae in Europe is driven by nosocomial spread. Nat Microbiol 4:1919–1929. https://doi.org/10.1038/s41564-019-0492-8.
24. Sherry NL, Lane CR, Kwong JC, Schultz M, Sait M, Stevens K, Ballard S, Goncalves da Silva A, Seemann T, Gorrie CL, Stinear TP, Williamson DA, Brett J, van Diemen A, Easton M, Howden BP. 2019. Genomics for molecular epidemiology and detecting transmission of carbapenemase-producing Enterobacterales in Victoria, Australia, 2012 to 2016. J Clin Microbiol 57:e00573-19.
25. Weingarten RA, Johnson RC, Conlan S, Ramsburg AM, Dekker JP, Lau AF, Khil P, Odom RT, Deming C, Park M, Thomas PJ, Program NCS, Henderson DK, Palmore TN, Segre JA, Frank KM. 2018. Genomic analysis of hospital plumbing reveals diverse reservoir of bacterial plasmids conferring carbapenem resistance. mBio 9:e02011-17. https://doi.org/10.1128/mBio.02011-17.
26. Cheruvanky A, Stoesser N, Sheppard AE, Crook DW, Hoffman PS, Weddle E, Carroll J, Sifri CD, Chai W, Barry K, Ramakrishnan G, Mathers AJ. 2017. Enhanced Klebsiella pneumoniae carbapenemase expression from a novel Tn4401 deletion. Antimicrob Agents Chemother 61:e00025-17. https://doi.org/10.1128/AAC.00025-17.
27. Bi D, Zheng J, Li JJ, Sheng ZK, Zhu X, Ou HY, Li Q, Wei Q. 2018. In silico typing and comparative genomic analysis of IncFIIK plasmids and insights into the evolution of replicons, plasmid backbones, and resistance determinant profiles. Antimicrob Agents Chemother 62:e00764-18. https://doi.org/10.1128/AAC.00764-18.
28. De Maio N, Shaw LP, Hubbard A, George S, Sanderson ND, Swann J, Wick R, AbuOun M, Stubberfield E, Hoosdally SJ, Crook DW, Peto TEA, Sheppard AE, Bailey MJ, Read DS, Anjum MF, Walker AS, Stoesser N, REHAB Consortium. 2019. Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. Microb Genom 5:e000294. https://doi.org/10.1099/mgen.0.000294.
29. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom 3:e000132. https://doi.org/10.1099/mgen.0.000132.
30. Lamble S, Batty E, Attar M, Buck D, Bowden R, Lunter G, Crook D, El-Fahmawi B, Piazza P. 2013. Improved workflows for high throughput library preparation using the transposome-based Nextera system. BMC Biotechnol 13:104. https://doi.org/10.1186/1472-6750-13-104.
31. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15:R46. https://doi.org/10.1186/gb-2014-15-3-r46.
32. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new

genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
33. Stoesser N, Batty EM, Eyre DW, Morgan M, Wyllie DH, Del Ojo Elias C, Johnson JR, Walker AS, Peto TE, Crook DW. 2013. Predicting antimicrobial susceptibilities for Escherichia coli and Klebsiella pneumoniae isolates using whole genomic sequence data. J Antimicrob Chemother 68:2234–2244. https://doi.org/10.1093/jac/dkt180.
34. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup F, Hasman H. 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother 58:3895–3903. https://doi.org/10.1128/AAC.02412-14.
35. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34:D32–D36. https://doi.org/10.1093/nar/gkj014.
36. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
37. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474.
38. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
39. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
40. Naas T, Cuzon G, Villegas MV, Lartigue MF, Quinn JP, Nordmann P. 2008. Genetic structures at the origin of acquisition of the beta-lactamase bla KPC gene. Antimicrob Agents Chemother 52:1257–1263. https://doi.org/10.1128/AAC.01451-07.
41. Villa L, Feudi C, Fortini D, Brisse S, Passet V, Bonura C, Endimiani A, Mammina C, Ocampo AM, Jimenez JN, Doumith M, Woodford N, Hopkins K, Carattoli A. 2017. Diversity, virulence, and antimicrobial resistance of the KPC-producing Klebsiella pneumoniae ST307 clone. Microbial Genomics 3:e000110. https://doi.org/10.1099/mgen.0.000110.
42. Orlek A, Phan H, Sheppard AE, Doumith M, Ellington M, Peto T, Crook D, Walker AS, Woodford N, Anjum MF, Stoesser N. 2017. A curated dataset of complete Enterobacteriaceae plasmids compiled from the NCBI nucleotide database. Data Brief 12:423–426. https://doi.org/10.1016/j.dib.2017.04.024.
43. Stoesser N, Sheppard AE, Peirano G, Anson LW, Pankhurst L, Sebra R, Phan HTT, Kasarskis A, Mathers AJ, Peto TEA, Bradford P, Motyl MR, Walker AS, Crook DW, Pitout JD. 2017. Genomic epidemiology of global Klebsiella pneumoniae carbapenemase (KPC)-producing Escherichia coli. Sci Rep 7:5917. https://doi.org/10.1038/s41598-017-06256-2.
44. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12. https://doi.org/10.1186/gb-2004-5-2-r12.
45. Frey BJ, Dueck D. 2007. Clustering by passing messages between data points. Science 315:972–976. https://doi.org/10.1126/science.1136800.
46. Kizny Gordon A, Phan HTT, Lipworth SI, Cheong E, Gottlieb T, George S, Peto TEA, Mathers AJ, Walker AS, Crook DW, Stoesser N. 2020. Genomic dynamics of species and mobile genetic elements in a prolonged blaIMP-4-associated carbapenemase outbreak in an Australian hospital. J Antimicrob Chemother 2020:dkz526. https://doi.org/10.1093/jac/dkz526.
47. Pritchard L, White JA, Birch PR, Toth IK. 2006. GenomeDiagram: a python package for the visualization of large-scale genomic data. Bioinformatics 22:616–617. https://doi.org/10.1093/bioinformatics/btk021.
48. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25:1422–1423. https://doi.org/10.1093/bioinformatics/btp163.